
序 文

分子レベルでの進化の研究は、2つの重要な問題を扱うことを目的としている。ひとつは種間の進化的な関係の再構築であり、もうひとつは進化過程の原動力や機構についての研究である。前者は分類学の領域であり、伝統的には形態的な形質や化石を用いて研究されている。しかし、分子データの有用性の高さと利用の容易さから、分子は、多くの生物種の系統の再構築に最も一般的に使われるデータのタイプになってきた。後者の、分子進化の機構に関わる問題は、塩基やアミノ酸の置換速度の推定や、配列データを使った突然変異や選択のモデルの検定によって研究されている。

急激な遺伝子配列データの蓄積、コンピュータのハードウェアとソフトウェアの性能の向上、興味深い生物学的な問題を扱うのにふさわしい精密な統計的手法の開発により、上記のどちらの分野の研究もこの数十年で急激に進展してきた。あらゆる兆候は、とりわけデータ生成の場において、この発展が今後も続くであろうことを示唆している。系統解析はゲノム時代に突入し、何百もの生物種や配列からなる大規模なデータセットが日常的に解析されている。“形態”対“分子”について論争はほぼ終わっている；大部分の研究者は、どちらのデータの価値も十分に認識している。最節約法と最尤法に関する哲学的な論争は続いているが、以前ほどのとげとげしさはない。はるかに刺激的な進展が、強力な統計的方法やモデルの開発や実装においてなされ、いまや実際のデータセットの解析に日常的に使用されている。

この分野における方法論的な進歩をまとめるのにふさわしい時期がきていると思われ、本書ではそれを試みている。しかし、本書では、分子進化研究の全体像をとらえようとはしていない。それについては、最近出版された Joseph Felsenstein (2004) の本で系統学に関するほぼすべての事柄が議論されているので、現時点ではそのような必要はほとんどないからだ。その代わりに、系統関係の再構築や進化過程の推定を含む分子進化解析は、統計的な推定の問題である (Cavalli-Sforza and Edwards, 1967) という視点から本書をまとめる。そのため、最尤法やベイズ法といった確立した統計手法を標準的なものとして取り扱う。発見的方法や近似法については統計的観点から記述し、またその単純さと直観的なわかりやすさから、より厳密な方法を説明する前に、中心となる概念を紹介するためにしばしば使用する。また、データ解析の方法を開発する研究者にとっても参考となりうるように、本書では実装の問題についても議論する。

本書は、上級の学部生、研究生、また進化生物学、分子分類学、集団遺伝学の研究者を対象としている。これまで自身のデータを解析するのにプログラム・ソフトウェアを使ってきた生物学

者が、その方法のはたらきを理解する上で本書が一助となることを期待する。この本では基本的な概念に力を入れているが、数学的な導出も詳しく解説しているので、この刺激的な計算生物学の分野をめざす統計学者、数学者、情報科学者が読んでもよいだろう。

本書は、たとえば Graur & Li (2000) の第 1 章にあるような基礎的な遺伝学の知識を読者がもちあわせていることを想定している。基本的な統計学あるいは生物統計学の知識も前提としており、微積分や線形代数もいくつかの章で必要となる。最尤法やベイズ統計は、簡単な例を使って導入し、その後でより複雑な解析に用いる。これらの手法について体系的に幅広く知りたい読者は、多くの確率論や数理統計学のすぐれた参考書を参照していただきたい。たとえば、初級者レベルには DeGroot and Schervish (2002) が、上級者レベルには Davison (2003), Stuart *et al.* (1999), Leonard & Hsu (1999) がある。

本書の構成は次のようになっている。第 I 部は 2 つの章からなり、配列進化のマルコフ過程モデルを紹介する。第 1 章では塩基置換のモデルや 2 つの配列間の距離の計算について議論する。これはおそらく最も簡単な系統解析である。また、本書の後半でひんばんに使うマルコフ連鎖の理論と最尤法について、ここで紹介する。そのため、この章は生物学者には最も努力を必要とする章かもしれない。第 2 章では、アミノ酸置換とコドン置換についてのマルコフ過程モデルと、2 つのタンパク質配列間の距離計算や、タンパク質をコードする 2 つの DNA 配列の同義置換率や非同義置換率の推定へのそれらのモデルの利用について述べる。第 II 部では系統関係の再構築の手法を取り扱う。最節約法や距離行列法については簡単に議論するが (第 3 章)、最尤法やベイズ法については深く掘り下げる (第 4 章, 第 5 章)。第 5 章は、Olivier Gascuel によって編集された *Mathematics in Phylogeny and Evolution* (Oxford University Press, 2005) 中で執筆した章の内容を拡張したものである。第 6 章は、いろいろな系統関係の再構築手法を比較した研究の総説となっており、系統樹の検定も含まれる。第 III 部では、進化過程を研究するための系統学的方法の応用のいくつかについて議論される。第 7 章では分子時計の検定や種分岐年代の推定のための分子時計の使用を、また第 8 章ではタンパク質の進化に影響を及ぼす自然選択を検出するためのコドン置換モデルの応用を議論する。第 9 章ではコンピュータ・シミュレーションの基本的な技法について議論する。第 10 章では、この分野における近年の動向や将来の展望が議論される。補遺 C では、系統解析の主要なソフトウェア・パッケージについて簡単に解説する。星印 (*) のついた項目は専門的であるので、読み飛ばしてもかまわない。

本書で使われた例題のデータセットや、本書で議論されているアルゴリズムを実装した C 言語による小さなプログラムは、本書に関するウェブサイトにおかれている：<http://abacus.gene.ucl.ac.uk/CME/>。そのサイトには、本書の出版後に見つかった誤りの一覧もある。誤りを見つけた場合には、E-mail: z.yang@ucl.ac.uk まで連絡していただきたい。

本書の各章の初期の原稿を読み、建設的なコメントや批評をいただいた多くの同僚に深く感謝したい；Hiroshi Akashi (第 2 章, 第 8 章), Adam Eyre-Walker (第 2 章), Jim Mallet (第 4 章, 第 5 章), Konrad Scheffler (第 1 章, 第 5 章, 第 8 章), Elliott Sober (第 6 章), Mike Steel (第 6 章), Jeff Thorne (第 6 章), Simon Whelan (第 1 章), Anne Yoder (第 6 章)。本書

全体にわたって目を通し，詳細な指摘をしてくれた Karen Cranston, Ligia Mateiu, Fengrong Ren にはとくに感謝したい。Jessica Vamathevan と Richard Emes は，表現のむずかしいくだりを検討しているときに，その検討につきあってもらった。いうまでもなく，すべての誤りは私に帰する。Oxford University Press の Ian Sherman と Stefanie Gehrig には，本書の企画からはじまり，執筆中を通しての的確な支援と忍耐力に感謝する。

Ziheng Yang (楊子恒)

London にて

2006 年 3 月